

SECRET

ANNEX B

INTELLIGENCE DATA BASES

SECRET

ANNEX B

Index of Tables

Table B-1	Sponsoring Organizations for Intelligence Data Bases	B - 6
Table B-2	Components Operating Intelligence Data Bases	B - 7
Table B-3	COINS Data Bases	B - 9
Table B-4	Data Base Descriptors for Positive Intelligence	B - 15
Table B-5	Subject Matter Descriptors for Intelligence Data Bases	B - 16
Table B-5 (Expanded)	Subject Matter Descriptors for Intelligence Data Bases (Tab 1)	B - 25

UNCLASSIFIED

ANNEX B

INTELLIGENCE DATA BASES

I. SUMMARY

- STAT 1.1 ☐ Number of Data Bases. The Intelligence Community has identified some 467 automated intelligence data bases upon which analysts place heavy reliance in the performance of their assigned activities. One hundred seventy of these data bases are available through COINS (Community On-Line Intelligence System) and DIAOLS (DIA On-Line System). Data can also be shared by means of the bulk transfer of information via mainline general communications networks, such as AUTODIN.
- STAT 1.2 ☐ Data Base Subject. Almost all of these data bases are classified, and a large number contain compartmented materials. There is a great diversity of subject matters, and topics are often further subdivided by particular geographic regions or countries. The scope and the level of subject matter detail in individual data bases invariably is a reflection of the particular scope of mission responsibilities of the civilian organization or the military command which the data base primarily serves. The military commands and their subordinate components are located worldwide.
- STAT 1.3 ☐ Varied Uses of Data Bases. Data base design also reflects the particular substantive functions into which the total intelligence business can be subdivided. For example, some data bases exist to

UNCLASSIFIED

help in the management and operation of intelligence collection activities. Others are oriented to intelligence sensors (e.g., Imagery, COMINT, ELINT, Telemetry, etc.) and they support the processing of raw intelligence information. Others relate to mapping, charting and geodesy; to counterintelligence; to investigative activities; and to one or another form of general support activity (e.g., personnel management, financial management, logistics, communications, etc.). It is, however, the data bases that are created to support the Production function that are of broadest interest to substantive intelligence analysts and their customers. Sharing these data bases has been underway for years in some cases, and further collaboration and shared use has been the subject of intensive interagency discussion. Cost/effective opportunities for further sharing will be examined on a case by case basis in the coming year.

STAT 1.4 ☐ Varied Formats. The internal organization and formatting of data bases is invariably a reflection of the particular manner in which the data are to be used by intelligence analysts or provided to external customers in the form of intelligence outputs. Some data bases have an orientation to support in-depth research. Others are intended to serve analysts and customers in time-limited situations, where indications, warning, information fusion and crisis management are the driving considerations.

STAT 1.5 ☐ Need for Coordination. The Intelligence Community as a whole has not established or enforced, and does not now possess, a common

set of rules and standing agreements for the creation, maintenance or sharing of intelligence data bases. Rather, individual organizations have perceived needs of their own, and have set about to fulfill them by creating organized bodies of information on particular subjects. If that information was found to be useful to other organizations, and security and need-to-know considerations could be accommodated, that data base was shared externally. If it became particularly popular over time, efforts were made to place it on DIAOLS-COINS. In this connection, a major interagency study, and a follow-up action program, was carried out in 1972-74 to identify additional data bases that might be put on the COINS network and to give further impetus to the use of that network as a Community vehicle. As a result, COINS was upgraded by DCI order to include compartmented data bases, and numerous changes and additions to the COINS-accessible data bases were accomplished.

- STAT 1.6 ☐ Data Base Sharing. In his annual guidance to the Community, promulgated in early 1977, the DCI noted the need for the Community to take organized and authoritative action that would lead to identifying certain much-used data bases as "community property." Attendant on this concept is the necessity to make an official assignment of responsibility to a designated organization to accomplish the maintenance and to ensure the timeliness and accuracy of the data base. In addition, the designated organization needs assurance that appropriate budgetary resources will be provided to carry out the assigned task. As the concept of distributed production responsibility becomes more

widespread in practice, formalized agreements among Community members on this subject must inevitably be reached, and performance thereunder must be monitored on behalf of the entire Community.

STAT 1.7 ☐ System Analyses of Data Base Alternatives. Inherent in any overall program to establish effective Community collaboration in the creation and shared use, as appropriate, of data bases is the necessity to examine trade-offs between automated, semi-automated and non-automated data bases. In each case, a further consideration is the validated and confirmed urgency of the need for a particular body of information to be shared among particular organizations. The technical alternatives to accomplish this result need to be appraised in cost/effectiveness terms. A current example of this situation is the CIA AEGIS (bibliographic index) data base, which Community members have requested be shared more broadly. The DCI's Intelligence Information Handling Committee (IHC) has a working group now looking into this issue.

STAT 1.8 ☐ Data Base Costs. A deficiency in data base management at this time is the difficulty experienced in obtaining reliable information on the costs of preparing and maintaining data bases, both automated and non-automated. This is a difficulty that is not unique to the Intelligence Community, and the OMB has pointed out that this is a government-wide phenomenon. Part of the difficulty is the semantic one of reaching interagency consensus on the kinds of people whose work

UNCLASSIFIED

and salary costs shall be included in tabulation of "ADP-related personnel." In spite of the lack of government-wide standards on this point, it is clear that the several types of costs inherent in data base creation, operation and maintenance should be given attention. The future work program of the Intelligence Community will examine this problem.

- STAT 1.9 ☐ Department/Agency Submits on Data Bases. This annex contains a preliminary analysis of the information on data bases which has been collected as a result of a recent community-wide data call. This analysis was constrained because the quality of data received was not homogeneous nor complete. The data are, however, useful as an indicator of the characteristics of this subject. The individual submissions are included in full in Annex F.

- STAT 1.10 ☐ The Next Step in Data Base Analysis. The next step is to involve all Community organizations in improving their original contributions and engaging in collaborative review and analyses of these materials. A specific task for the Community's strengthened central planning structure (see Annex C) will be to recomplete these data in a more detailed and orderly manner (see Annex E discussion of a management information system). This work program is scheduled for 1978.

UNCLASSIFIED

STAT

1.11 ☐ Data Base Sponsoring Organizations. A tabulation of data bases reported on herein appears below and identifies the number of data bases which each member of the Intelligence Community has sponsored. Because of the variations in the manner of reporting the underlying information, the figures shown in Table B-1 should be taken as approximations. In particular, no effort was made to accommodate the size of a particular data base. Therefore, any particular inter-agency comparison of the data in the table is superfluous. /

(U) Table B-1
SPONSORING ORGANIZATIONS FOR
INTELLIGENCE DATA BASES

Army	81
Navy	126
Air Force	88
NSA	29
DIA	76
CIA	57
Intelligence Community Staff	5
FBI	4
DOE	1
State/INR	0
Treasury	0
Total	467

25X1A

Approved For Release 2002/09/05 : CIA-RDP84-00933R000200010005-6

Next 3 Page(s) In Document Exempt

Approved For Release 2002/09/05 : CIA-RDP84-00933R000200010005-6

STAT

2.3 ☐ Descriptors for Data Bases and Files. (*) The following tabulations herein present a number of orderly sets of descriptors -- that is, words which, when applied to intelligence data bases and files, make it possible to group them in general categories. Each of the descriptors that appears below is applicable to one or more of the files which has (have) been reported by Intelligence Community members in connection with the data call for this report. These descriptors are useful to categorize the numerous different functional aspects of the total intelligence business, using terminology that is more or less in general use throughout the Community.

(*) - In this discussion, a file is a grouping together of information on a defined topic, and a data base is simply a large file or a combination of files that make(s) up a more voluminous, orderly mass of information on a defined topic. Both files and data bases may be fully-, semi- or non-automated.

- In intelligence training manuals, it is customary to speak of information as that which is of interest to some aspect of the intelligence business and which is therefore collected. After it has been tested, analyzed, compared, and subjected to other appropriate processing, it comes to be called intelligence.

- Increasingly, the word data is used to mean raw material out of which information can be created. This is particularly the case when the so-called data are non-verbal (e.g., analog signals). Data is also used in casual discussion where its precise meaning must be derived from the context of its use.

UNCLASSIFIED

STAT

- 2.4 ☐ Standard Terminology. The terminology used in this annex is largely drawn from that which has been developed and officially prescribed for the Community Intelligence Resources Information System (CIRIS), a community-wide management mechanism and data base that categorizes in several sets of descriptors (e.g., mission, function, sensor, platform, target, etc.) the planned application by Intelligence components (called Reporting Entities) of their resources (dollars and manpower). The CIRIS data are tied directly to official control documents dealing with resources, such as the DoD Five Year Defense Plan (FYDP), and resource data of non-Defense department and agencies.

STAT

- 2.5 ☐ Methodology for Data Base Analysis. In the analysis leading to the preparation of this annex, a tabulation was laid out in working draft in the form of an extremely large matrix. In concept, this matrix makes it possible to identify all of the 467 data bases and all of their sponsors (Tables B-1 and B-2) on one axis. The matrix provides for listing all of the descriptive terminology (Tables B-4 and B-5) on the other axis. At the intersections, annotations can be entered to indicate the characteristics of the data bases by means of brief comments. Because of its great size as well as the uneven and tentative quality of some of the data reported, this matrix is not reproduced for this report.

STAT

- 2.6 ☐ Completing the Matrix. The completion of this matrix, and its further analysis, will provide a first-cut working tool to highlight

UNCLASSIFIED

UNCLASSIFIED

relationships among data bases. It can point to partial similarities among data bases as well as to their differences. It can serve as a general road map to identify relationships that need to be looked into with greater precision and in greater depth. The matrix clearly does not allow the making of definitive judgments solely on that evidence, and no attempt has been made herein to do so. The evidence presented is not now adequate, and considerable additional time will be required to do this in-depth analysis. The matrix does present an agenda for follow-on action at a technical working level.

STAT 2.7 ☐ Data Base Descriptors. The orderly sets of terms (taxonomies) that can be used to describe the Intelligence Community's data bases as reported on herein are presented in Tables B-4 and B-5. This listing is revealing because it illustrates the great diversity of subjects that are contained in these data bases, and it given an indication of the diverse uses to which these organized bodies of information are put.

STAT 2.8 ☐ Tables B-4 - Position Intelligence Descriptors. Consistent with the CIRIS structure, the total resources of the Intelligence Community can be subdivided initially into three types of work

B-13

UNCLASSIFIED

UNCLASSIFIED

called "missions." These are the Positive Intelligence mission, the Counterintelligence & Investigative Activities mission, and the General Support mission. Some of the data bases reported on herein assist the work of each of these broad missions. The largest part of the intelligence budget, and the largest number of these data bases, however, relate to the Positive Intelligence mission. Within Positive Intelligence, the work carried out can be further categorized by "functions." The purposes served by performing these functions can be grouped as "uses" of data bases. The manner in which the work of the Collection and the Processing Functions is accomplished is by means of "sensors." Table B-4 lists the terminology applicable to the uses, functions and sensors of the Positive Intelligence mission.

UNCLASSIFIED

25X1A

Approved For Release 2002/09/05 : CIA-RDP84-00933R000200010005-6

Approved For Release 2002/09/05 : CIA-RDP84-00933R000200010005-6

UNCLASSIFIED

STAT

2.9 ☐ Table B-5 - Subject Matter Descriptors. Table B-5 contains descriptors that characterize the subjects addressed by the data bases reported on herein. These subjects are consistent with the CIRIS definitions. Each topical heading in Table B-5 can be further broken down into sub-topics, each of which can be found in one or more of the data bases covered by this report. This further level of detail is presented at Tab 1.

(U) Table B-5

SUBJECT MATTER DESCRIPTORS
for
INTELLIGENCE DATA BASES

Mission:

Topical Heading:

Positive
Foreign
Intelligence

Political
Economic
Science & Technology
Physical Environment
Telecommunications
Military - Ground
Military - Sea
Military - Air and Space
Military - Order of Battle
Military - General
Politico-Military
Biographic
Reference Services Data Bases
Other:
(e.g. Individual Project Data Base)

Counterintel
&
Investig Act'ys

Counterintelligence Subjects
Investigative Activities Subjects

General
Support

General Support Subjects:
(e.g. Personnel, Security, Logistics,
ADP Management, General Administration)

UNCLASSIFIED

STAT

2.10 ☐ Network Access to Data Bases. Intelligence data bases can also be characterized by their availability via a network, such as COINS or DIAOLS, or their non-availability except via hand or mail delivery or by telephone inquiry.

STAT

2.11 ☐ Security Aspects. Intelligence data bases generally contain classified materials. Any analysis that deals with expanding access to data bases must take account of complex and presently unresolved matters of security policy. Progress here is a prerequisite to implementing technical solutions to the multi-level security problem. Issues in this field, accordingly, extend beyond the area of authority of information system managers, and they involve both security specialists and top management policy makers. The need for new solutions to the multi-level security problem extends beyond the Intelligence Community and is shared in one form or another by the entire government. Intelligence ADP managers uniformly report that they face no more pressing problem than this one. (This topic is addressed at greater length in the basic report, Section III.C.) Intelligence data bases can be categorized by their security characteristics, as follows: Unclassified, Confidential, Secret, Top Secret, SI Compartment, TK Compartment, Other Compartment(s), and No Foreign Dissemination (NOFORN). In addition, there are other controls that may be specified for intelligence data bases. A separate but important problem is the dissemination-limiting controls that exist outside the Intelligence Community, and that may be imposed by

UNCLASSIFIED

officials planning or conducting military operations and foreign policy activities. This is the Operations-Intelligence interface problem. It impacts both on the building of intelligence data bases and on their use, particularly in crisis situations.

III. OBSERVATIONS

- STAT 3.1 ☐ Introduction. The information relative to data bases and their improvement collected for this report is adequate for a general characterization of this subject, such as has been presented in this annex. A large amount of future work will be required to plan for the cost cost/effective future development and use of intelligence data bases and to implement that planning. The following observations will provide some general guidance and a focus for future work in this area.
- STAT 3.2 ☐ Intelligence Work is a Continuum. Intelligence work reflects a continuing flow of activity -- from the collector, through the processor of that which has been collected, to the substantive analyst who is responsible for selecting pertinent information, performing analytical manipulations, and synthesizing the materials in order to reach the ultimate result which is the completion of a finished product.
- STAT 3.3 ☐ Collection and Processing Data Bases are Specialized. Collectors and processors of information gathered through technical means specialize in a particular kind of intelligence (e.g., a communications

UNCLASSIFIED

signal, a picture, or data derived from specialized instrumentation). The working data bases of collectors and processors are designed and oriented to help them carry out their own tasks. As such, these data bases are not useful ordinarily to customers outside the Intelligence Community, nor are they particularly useful to intelligence analysts in general. The outputs from some processors, such as the imagery product from the National Photographic Interpretation Center (NIPC) and the SIGINT product from NSA, are very important to intelligence analysts, and in these cases arrangements already exist to share these intelligence materials across agency lines via the DIAOLS-COINS network.

- STAT 3.4 ☐ Production Data Bases are of Broader Interest. It is the data bases that are created to support the Production function that are of broadest interest to intelligence analysts and their consumers. There is, for example, a large family of military intelligence data bases that involves order of battle information, and another family that involves foreign installations. The detailed information presented in Annex F demonstrates that as between DIA, the Armed Services, and the U&S Commands, there is much activity underway to share these kinds of data bases. Much of this kind of information moves via bulk transfer over Defense telecommunications networks. Procedures have been instituted by DIA to involve data base managers and users in an overall plan for sharing these data bases. DIA, for example, maintains a number of master data files for the DoD community and

through DIAOLS provides for distribution or on-line access to more than 1200 users. Other organizations have planning underway or existing procedures to permit wider access to their files. A few illustrations of the trend to greater interconnectivity of data base sharing in GDIP include the following: the ASSIST and EUCOM AIDES programs within the Army, the CIRC program by FTD in the Air Force for S&T materials, the Navy's OSIS, the expanded uses of the SAC PACER data base.

STAT 3.5 ☐ Central Reference Services. Table 1, which expands on Table B-5, lists a family of files that pertain to the contents of central reference services and to their administration and operation. As has been noted above, the CIA's bibliographic index, (AEGIS), will be studied during 1978 to determine if community-wide access should occur. Another topic which warrants further examination is the handling and filing of electrical communications, and this is likewise an agenda item for 1978. Today, the information handling role of central reference services extends far beyond the obvious activities of the traditional library. Much intelligence data is message and report oriented, and central reference services provide important forms of substantive support, working in partnership with

B - 20

UNCLASSIFIED

production analysts. Increasingly, reference service personnel provide an informed human interface between an external, automated data base and an in-house analyst who is not familiar with the procedures to access remote data bases via a computer terminal. Central reference service personnel are experts also in the growing field of microforms, as increasing amounts of intelligence are stored in this medium. As the quantity of information continues to proliferate, and as the complexity of intelligence analysis increases, central reference services and their staffs play roles of critical importance in mastering the data and serving their customers.

- STAT 3.6 ☐ Analyst Working Files. One very large category of files, regardless of subject matter, can be described as "substantive analyst working files." This kind of files is not reported on in Annex F, nor should it be. It is the materials that exist in working form in the analyst's local office file cabinets and sometimes on a local automated system. Project SAFE, for example, seeks to give CIA and DIA analysts a greater capability to handle, manipulate and set up working files of selected materials that pertain to their own areas of expertise and responsibility. It is only when an organized body of knowledge, whether called a file or a data base, has reached such a stage of maturity and stability that it is seen as a central reference service asset that it is ready for evaluation as to its potential for being shared beyond the local organization that created it as an unofficial working tool.

STAT

3.7

☐ Formatting and Standardization of Data Bases. Some files,

by virtue of the subject matter, lend themselves to a considerable degree of internal organization, called formatting. Military force order-of-battle files are of this character. Other files, however, may address a multi-subject problem that is not capable of being tightly described or structured, and in this case formatting is not feasible. Another more subtle limitation on formatting is in the emphasis of the using intelligence analyst: what one analyst may find of crucial importance may be only of slight concern to another analyst dealing with the same general subject matter but from a different point of view and for a totally different customer and use. In consequence, greater formatting and standardization of files may, upon precise evaluation, prove to be both costly and unsatisfactory. The "data element standardization" issue is discussed in the basic report in Section III. C., and this discussion points out that the question of whether or not data element standardization is cost/effective will depend on a specific analysis of each file and its particular uses. The Community needs to address a number of these specific cases in the immediate future.

STAT

3.8

☐ Analyst Needs Should Control Data Base Design. A very im-

portant consideration to guide the future work of examining alternatives to enhance the quality, accessibility, shareability, and timeliness of Intelligence Community files and data bases is a thorough appreciation of the role and requirements of the intelligence

analyst who is to benefit from the use of these materials. It has been demonstrated beyond argument, both within the government and outside, that an analyst will not learn to use an ADP file if the process of learning is too complex and hence over-demanding on his limited stock of time. Nor will the analyst give up his paper files and depend totally on an automated file until he has gained complete confidence that the automated system will not let him down -- that it will be accurate and that it will be available promptly whenever he needs it, particularly in pressing, time-limited situations. Automated systems that manage intelligence data bases must, therefore, be built to an extremely high standard of reliability. Since this level of assurance can be expensive, future analyses of the role of automation need to look closely at the trade-offs between automated and non-automated systems and procedures, and between centralized automated systems and local automated aids to the analyst that he can use without the possible complexity of being involved in a large centrally-operated system. The problems vary and so do the answers.

IV. SUMMARY

- STAT 4.1 ☐ This annex has identified the sets of descriptors that apply to the intelligence data bases reported on herein. A general characterization of those data bases, has been presented, and it has suggested that a more thorough examination should be undertaken

through collaborative efforts of the Community during 1978. It has suggested that the future focus be placed particularly, but not exclusively, on Production-function-oriented data bases.

25X1

Approved For Release 2002/09/05 : CIA-RDP84-00933R000200010005-6

Next 2 Page(s) In Document Exempt

Approved For Release 2002/09/05 : CIA-RDP84-00933R000200010005-6